

Homework 5

Autumn 2022

YOUR NAME

Instructions

- This homework is due in Gradescope on Wednesday Nov 9 by midnight PST.
 - Please answer the following questions in the order in which they are posed.
 - Don't forget to knit the document frequently to make sure there are no compilation errors.
 - When you are done, download the PDF file as instructed in section and submit it in Gradescope.
-

1. (Linear transformations) Suppose that the PMF of X is given by:

x	a	b
$f(x)$	π	$(1 - \pi)$

where a and b are some numbers. Answer the following. Be sure to support your answers.

- Find $E[X]$.
 - Find $Var[X]$.
 - Find $E\left[\frac{X-b}{a-b}\right]$.
 - Find $Var\left[\frac{X-b}{a-b}\right]$.
2. (Rh negative) In the US population, 85% have an agglutinating factor in their blood classifying them as Rh positive, while 15% lack the factor and are Rh negative. A medical researcher wants to analyze blood from a newborn Rh negative infant, so he examines the blood types of successive newborn infants until he finds an Rh negative infant.
- How many Rh positive infants should they expect to type before they find their first Rh negative? Be sure to set up the problem (random variable, distribution, assumptions you need to make) and clearly **state** any results you are using before just using them.
 - Calculate the probability that they will type more Rh positive infants than expected. This probability can be calculated exactly as shown in class. Give the closed form expression for this probability and then calculate it in R. Be sure to echo your code chunk.
3. (Memoryless) Let X be a geometric random variable with probability π . That is,

$$f(x) = \pi \cdot (1 - \pi)^x, \quad x = 0, 1, 2, \dots$$

- Let k be some non-negative integer. What is $P(X \geq k)$? (*Hint*: we did this in class, I want you to do it again.)

- b. Show that for all non-negative integers $x(\geq k)$

$$P(X \geq x | X \geq k) = P(X \geq x - k).$$

- c. Because of this result, we say that the geometric distribution is *memoryless*. Explain how this is an appropriate name for this property.
4. (Apple trees) From each of 6 trees in an apple orchard, 25 leaves are selected. On each of the leaves, the number of adult mites are counted.

x	0	1	2	3	4	5	6	7	8+
count	70	38	17	10	9	3	2	1	0

The dataset can be created in R as shown below. Type `?rep` for help.

- a. Print the number of observations (leaves) in the data set along with the mean \bar{x} and standard deviation s of the `mites` variable. Be sure to show your code.

Let X denote the number of mites on a randomly selected leaf. In this question, you will consider two possible models for X :

```
- $X ~sim Poisson(\lambda = \bar{x})$
- $X ~sim Geom( prob = \frac{1}{1 + \bar{x}} )$
```

where \bar{x} is the number you calculated from part a. for the mean.

- b. Make side-by-side plots showing how well each model fits. Each plot must have the histogram of the data with the probabilities according to the model overlaid on it. Which model appears to fit the data better? (Be sure to echo your code. Don't forget labels, titles etc.)

Hint: For help with the code for side-by-side plots, look at Section 8.1 on the negative binomial. For help on overlaying theoretical probabilities on a histogram, look at section 8.2 on the Poisson distribution.